CCGPS Frameworks Student Edition

Mathematics

CCGPS Coordinate Algebra

Unit 4: Describing Data

Dr. John D. Barge, State School Superintendent
"Making Education Work for All Georgians"

Unit 4
Describing Data

Table of Contents

OVERVIEW ... 3
STANDARDS ADDRESSED IN THIS UNIT ... 3
KEY & RELATED STANDARDS ... 3
ENDURING UNDERSTANDINGS ... 7
CONCEPTS AND SKILLS TO MAINTAIN ... 8
SELECT TERMS AND SYMBOLS ... 8
TASKS
- Math Class ... 11
- If the Shoe Fits! ... 12
- The Basketball Star ... 16
- Public Opinions ... 20
- Leisure Time ... 21
- Spaghetti Regression ... 22
- TV/Test Grades ... 25
- Equal Salaries for Equal Work? ... 27

OVERVIEW

In this unit student will:

- Assess how a model fits data
- Choose a summary statistic appropriate to the characteristics of the data distribution, such as the shape of the distribution or the existence of extreme data points
- Use regression techniques to describe approximately linear relationships between quantities.
- Use graphical representations and knowledge of the context to make judgments about the appropriateness of linear models
- Look at residuals to analyze the goodness of fit. Students take a more sophisticated look at using a linear function to model the relationship between two numerical variables.

Although the units in this instructional framework emphasize key standards and big ideas at specific times of the year, routine topics such as estimation, mental computation, and basic computation facts should be addressed on an ongoing basis. Ideas related to the eight practice standards should be addressed constantly as well. To assure that this unit is taught with the appropriate emphasis, depth, and rigor, it is important that the tasks listed under "Evidence of Learning" be reviewed early in the planning process. A variety of resources should be utilized to supplement this unit. This unit provides much needed content information, but excellent learning activities as well. The tasks in this unit illustrate the types of learning activities that should be utilized from a variety of sources.

STANDARDS ADDRESSED IN THIS UNIT

Mathematical standards are interwoven and should be addressed throughout the year in as many different units and activities as possible in order to emphasize the natural connections that exist among mathematical topics.

KEY & RELATED STANDARDS

Interpreting Categorical and Quantitative Data

Summarize, represent, and interpret data on a single count or measurement variable.

MCC9-12.S.ID.1 Represent data with plots on the real number line (dot plots, histograms, and box plots). Choose appropriate graphs to be consistent with numerical data: dot plots, histograms, and box plots.

MCC9-12.S.ID.2 Use statistics appropriate to the shape of the data distribution to compare center (median, mean) and spread (interquartile range, standard deviation-Advanced Algebra) of two or more different data sets. Include review of Mean Absolute Deviation as a measure of variation.

MCC9-12.S.ID.3 Interpret differences in shape, center, and spread in the context of the data sets, accounting for possible effects of extreme data points (outliers). Students will examine graphical representations to determine if data are symmetric, skewed left, or skewed right and how the shape of the data affects descriptive statistics.

Summarize, represent, and interpret data on two categorical and quantitative variables.

MCC9-12.S.ID.5 Summarize categorical data for two categories in two-way frequency tables. Interpret relative frequencies in the context of the data (including joint, marginal, and conditional relative frequencies). Recognize possible associations and trends in the data.

MCC9-12.S.ID.6 Represent data on two quantitative variables on a scatter plot, and describe how the variables are related.

MCC9-12.S.ID.6a Fit a function to the data; use functions fitted to data to solve problems in the context of the data. Use given functions or choose a function suggested by the context. Emphasize linear, ~~quadratic,~~ and exponential models.

MCC9-12.S.ID.6b Informally assess the fit of a function by plotting and analyzing residuals.

MCC9-12.S.ID.6c Fit a linear function for a scatter plot that suggests a linear association.

Interpret linear models

MCC9-12.S.ID.7 Interpret the slope (rate of change) and the intercept (constant term) of a linear model in the context of the data.

MCC9-12.S.ID.8 Compute (using technology) and interpret the correlation coefficient of a linear fit.

MCC9-12.S.ID.9 Distinguish between correlation and causation.

RELATED STANDARD

MCC6.SP.5 Summarize numerical data sets in relation to their context, such as by:
c. Giving quantitative measures of center (median and/or mean) and variability (interquartile range and/or mean absolute deviation), as well as describing any overall pattern and any striking deviations from the overall pattern with reference to the context in which the data was gathered.

Standards for Mathematical Practice

The Standards for Mathematical Practice describe varieties of expertise that mathematics educators at all levels should seek to develop in their students. These practices rest on important “processes and proficiencies” with longstanding importance in mathematics education. The first of these are

the NCTM process standards of problem solving, reasoning and proof, communication, representation, and connections. The second are the strands of mathematical proficiency specified in the National Research Council's report *Adding It Up*: adaptive reasoning, strategic competence, conceptual understanding (comprehension of mathematical concepts, operations and relations), procedural fluency (skill in carrying out procedures flexibly, accurately, efficiently and appropriately), and productive disposition (habitual inclination to see mathematics as sensible, useful, and worthwhile, coupled with a belief in diligence and one's own efficacy).

1. **Make sense of problems and persevere in solving them.** High school students start to examine problems by explaining to themselves the meaning of a problem and looking for entry points to its solution. They analyze givens, constraints, relationships, and goals. They make conjectures about the form and meaning of the solution and plan a solution pathway rather than simply jumping into a solution attempt. They consider analogous problems, and try special cases and simpler forms of the original problem in order to gain insight into its solution. They monitor and evaluate their progress and change course if necessary. Older students might, depending on the context of the problem, transform algebraic expressions or change the viewing window on their graphing calculator to get the information they need. By high school, students can explain correspondences between equations, verbal descriptions, tables, and graphs or draw diagrams of important features and relationships, graph data, and search for regularity or trends. They check their answers to problems using different methods and continually ask themselves, "Does this make sense?" They can understand the approaches of others to solving complex problems and identify correspondences between different approaches.

2. **Reason abstractly and quantitatively.** High school students seek to make sense of quantities and their relationships in problem situations. They abstract a given situation and represent it symbolically, manipulate the representing symbols, and pause as needed during the manipulation process in order to probe into the referents for the symbols involved. Students use quantitative reasoning to create coherent representations of the problem at hand; consider the units involved; attend to the meaning of quantities, not just how to compute them; and know and flexibly use different properties of operations and objects.

3. **Construct viable arguments and critique the reasoning of others.** High school students understand and use stated assumptions, definitions, and previously established results in constructing arguments. They make conjectures and build a logical progression of statements to explore the truth of their conjectures. They are able to analyze situations by breaking them into cases, and can recognize and use counterexamples. They justify their conclusions, communicate them to others, and respond to the arguments of others. They reason inductively about data, making plausible arguments that take into account the context from which the data arose. High school students are also able to compare the effectiveness of two plausible arguments, distinguish correct logic or reasoning from that which is flawed, and—if there is a flaw in an argument—explain what it is. High school students learn to determine domains to which an argument applies, listen or read the arguments of others, decide whether they make sense, and ask useful questions to clarify or improve the arguments.

4. **Model with mathematics.** High school students can apply the mathematics they know to solve problems arising in everyday life, society, and the workplace. By high school, a student might use geometry to solve a design problem or use a function to describe how one quantity of interest depends on another. High school students making assumptions and approximations to simplify a complicated situation, realizing that these may need revision later. They are able to identify important quantities in a practical situation and map their relationships using such tools as diagrams, two-way tables, graphs, flowcharts and formulas. They can analyze those relationships mathematically to draw conclusions. They routinely interpret their mathematical results in the context of the situation and reflect on whether the results make sense, possibly improving the model if it has not served its purpose.

5. **Use appropriate tools strategically.** High school students consider the available tools when solving a mathematical problem. These tools might include pencil and paper, concrete models, a ruler, a protractor, a calculator, a spreadsheet, a computer algebra system, a statistical package, or dynamic geometry software. High school students should be sufficiently familiar with tools appropriate for their grade or course to make sound decisions about when each of these tools might be helpful, recognizing both the insight to be gained and their limitations. For example, high school students analyze graphs of functions and solutions generated using a graphing calculator. They detect possible errors by strategically using estimation and other mathematical knowledge. When making mathematical models, they know that technology can enable them to visualize the results of varying assumptions, explore consequences, and compare predictions with data. They are able to identify relevant external mathematical resources, such as digital content located on a website, and use them to pose or solve problems. They are able to use technological tools to explore and deepen their understanding of concepts.

6. **Attend to precision.** High school students try to communicate precisely to others by using clear definitions in discussion with others and in their own reasoning. They state the meaning of the symbols they choose, specifying units of measure, and labeling axes to clarify the correspondence with quantities in a problem. They calculate accurately and efficiently, express numerical answers with a degree of precision appropriate for the problem context. By the time they reach high school they have learned to examine claims and make explicit use of definitions.

7. **Look for and make use of structure.** By high school, students look closely to discern a pattern or structure. In the expression $x^2 + 9x + 14$, older students can see the 14 as 2×7 and the 9 as $2 + 7$. They recognize the significance of an existing line in a geometric figure and can use the strategy of drawing an auxiliary line for solving problems. They also can step back for an overview and shift perspective. They can see complicated things, such as some algebraic expressions, as single objects or as being composed of several objects. For example, they can see $5 - 3(x - y)^2$ as 5 minus a positive number times a square and use that to realize that its value cannot be more than 5 for any real numbers x and y. High

school students use these patterns to create equivalent expressions, factor and solve equations, and compose functions, and transform figures.

8. **Look for and express regularity in repeated reasoning.** High school students notice if calculations are repeated, and look both for general methods and for shortcuts. Noticing the regularity in the way terms cancel when expanding $(x – 1)(x + 1)$, $(x – 1)(x^2 + x + 1)$, and $(x – 1)(x^3 + x^2 + x + 1)$ might lead them to the general formula for the sum of a geometric series. As they work to solve a problem, derive formulas or make generalizations, high school students maintain oversight of the process, while attending to the details. They continually evaluate the reasonableness of their intermediate results.

Connecting the Standards for Mathematical Practice to the Standards for Mathematical Content

The Standards for Mathematical Practice describe ways in which developing student practitioners of the discipline of mathematics should engage with the subject matter as they grow in mathematical maturity and expertise throughout the elementary, middle and high school years. Designers of curricula, assessments, and professional development should all attend to the need to connect the mathematical practices to mathematical content in mathematics instruction.

The Standards for Mathematical Content are a balanced combination of procedure and understanding. Expectations that begin with the word “understand” are often especially good opportunities to connect the practices to the content. Students who do not have an understanding of a topic may rely on procedures too heavily. Without a flexible base from which to work, they may be less likely to consider analogous problems, represent problems coherently, justify conclusions, apply the mathematics to practical situations, use technology mindfully to work with the mathematics, explain the mathematics accurately to other students, step back for an overview, or deviate from a known procedure to find a shortcut. In short, a missing mathematical knowledge effectively prevents a student from engaging in the mathematical practices.

In this respect, those content standards which set an expectation of understanding are potential “points of intersection” between the Standards for Mathematical Content and the Standards for Mathematical Practice. These points of intersection are intended to be weighted toward central and generative concepts in the school mathematics curriculum that most merit the time, resources, innovative energies, and focus necessary to qualitatively improve the curriculum, instruction, assessment, professional development, and student achievement in mathematics.

ENDURING UNDERSTANDINGS

- Data are gathered, displayed, summarized, examined, and interpreted to discover patterns and deviations from patterns.

- Which statistics to compare, which plots to use, and what the results of a comparison might mean, depend on the question to be investigated and the real-life actions to be taken.

- Understand and be able to use the context of the data to explain why its distribution takes on a particular shape (e.g. are there real-life limits to the values of the data that force skewness?)

- When making statistical models, technology is valuable for varying assumptions, exploring consequences and comparing predictions with data.

- Causation implies correlation yet correlation does not imply causation.

CONCEPTS AND SKILLS TO MAINTAIN

In order for students to be successful, the following skills and concepts need to be maintained

- Know how to compute the mean, median, interquartile range, and mean standard deviation by hand in simple cases and using technology with larger data sets.
- Find the lower extreme (minimum), upper extreme (maximum), and quartiles.
- Create a graphical representation of a data set.
- Present data in a frequency table.
- Plot data on a coordinate grid and graph linear functions.
- Recognize characteristics of linear and exponential functions.
- Write an equation of a line given two points.
- Graph data in a scatter plot and determine a trend.
- Determine the slope of a line from any representation.
- Identify the y intercept from any representation.
- Be able to use graphing technology.
- Understand the meaning of correlation.

SELECT TERMS AND SYMBOLS

The following terms and symbols are often misunderstood. These concepts are not an inclusive list and should not be taught in isolation. However, due to evidence of frequent difficulty and misunderstanding associated with these concepts, instructors should pay particular attention to them and how their students are able to explain and apply them.

The definitions below are for teacher reference only and are not to be memorized by the students. Students should explore these concepts using models and real life examples. Students should understand the concepts involved and be able to recognize and/or demonstrate them with words, models, pictures, or numbers.

The websites below are interactive and include a math glossary suitable for high school children. **Note – At the high school level, different sources use different definitions. Please preview any website for alignment to the definitions given in the frameworks**
http://www.amathsdictionaryforkids.com/
This web site has activities to help students more fully understand and retain new vocabulary.
http://intermath.coe.uga.edu/dictnary/homepg.asp

Definitions and activities for these and other terms can be found on the Intermath website. Intermath is geared towards middle and high school students.

- **Association.** A connection between data values.

- **Bivariate data.** Pairs of linked numerical observations. Example: a list of heights and weights for each player on a football team.

- **Box Plot.** A method of visually displaying a distribution of data values by using the median, quartiles, and extremes of the data set. A box shows the middle 50% of the data.

- **Box-and-Whisker Plot.** A diagram that shows the five-number summary of a distribution. (Five-number summary includes the minimum, lower quartile (25^{th} percentile), median (50^{th} percentile), upper quartile (75^{th} percentile), and the maximum. In a modified box plot, the presence of outliers can also be illustrated.

- **Categorical Variables.** Categorical variables take on values that are names or labels. The color of a ball (e.g., red, green, blue), gender (male or female), year in school (freshmen, sophomore, junior, senior). These are data that cannot be averaged or represented by a scatter plot as they have no numerical meaning.

- **Center.** Measures of center refer to the summary measures used to describe the most "typical" value in a set of data. The two most common measures of center are median and the mean.

- **Conditional Frequencies.** The relative frequencies in the body of a two-way frequency table.

- **Correlation Coefficient.** A measure of the strength of the linear relationship between two variables that is defined in terms of the (sample) covariance of the variables divided by their (sample) standard deviations.

- **Dot plot.** A method of visually displaying a distribution of data values where each data value is shown as a dot or mark above a number line.

- **First Quartile (**Q_1). The "middle value" in the *lower* half of the rank-ordered data

- **Histogram-** Graphical display that subdivides the data into class intervals and uses a rectangle to show the frequency of observations in those intervals—for example you might do intervals of 0-3, 4-7, 8-11, and 12-15

- **Interquartile Range.** A measure of variation in a set of numerical data. The interquartile range is the distance between the first and third quartiles of the data set. Example: For the data set {1, 3, 6, 7, 10, 12, 14, 15, 22, 120}, the interquartile range is 15 – 6 = 9.

- **Joint Frequencies.** Entries in the body of a two-way frequency table.

- **Line of best fit (**trend or regression line). A straight line that best represents the data on a scatter plot. This line may pass through some of the points, none of the points, or all of the points. Remind students that an exponential model will produce a curved fit.

- **Marginal Frequencies.** Entries in the "Total" row and "Total" column of a two-way frequency table.

- **Mean absolute deviation.** A measure of variation in a set of numerical data, computed by adding the distances between each data value and the mean, then dividing by the number of data values. Example: For the data set {2, 3, 6, 7, 10, 12, 14, 15, 22, 120}, the mean absolute deviation is 20.

- **Outlier.** Sometimes, distributions are characterized by extreme values that differ greatly from the other observations. These extreme values are called outliers. As a rule, an extreme value is considered to be an outlier if it is at least 1.5 interquartile ranges below the lower quartile (Q_1), or at least 1.5 interquartile ranges above the upper quartile (Q_3).

  OUTLIER if the values lie outside these specific ranges:

  $Q_1 - 1.5 \bullet IQR$

  $Q_3 + 1.5 \bullet IQR$

- **Quantitative Variables.** Numerical variables that represent a measurable quantity. For example, when we speak of the population of a city, we are talking about the number of people in the city – a measurable attribute of the city. Therefore, population would be a quantitative variable. Other examples: scores on a set of tests, height and weight, temperature at the top of each hour.

- **Residuals** (error). Represents unexplained (or residual) variation after fitting a regression model. **residual** = observed value – predicted value $e = y - \hat{y}$. A **residual plot** is a graph that shows the residual values on the vertical axis and the independent (*x*) variable on the horizontal axis.

- **Scatter plot.** A graph in the coordinate plane representing a set of bivariate data. For example, the heights and weights of a group of people could be displayed on a scatter plot. If you are looking for values that fall within the range of values plotted on the scatter plot, you are interpolating. If you are looking for values that fall beyond the range of those values plotted on the scatter plot, you are extrapolating.

- **Second Quartile (**Q_2**).** The *median* value in the data set.

- **Shape**. The shape of a distribution is described by symmetry, number of peaks, direction of skew, or uniformity.

- **Symmetry**- A symmetric distribution can be divided at the center so that each half is a mirror image of the other.
- **Number of Peaks**- Distributions can have few or many peaks. Distributions with one clear peak are called unimodal and distributions with two clear peaks are called bimodal. Unimodal distributions are sometimes called bell-shaped.
- **Direction of Skew**- Some distributions have many more observations on one side of graph than the other. Distributions with a tail on the right toward the higher values are said to be skewed right; and distributions with a tail on the left toward the lower values are said to be skewed left.
- **Uniformity-** When observations in a set of data are equally spread across the range of the distribution, the distribution is called uniform distribution. A uniform distribution has no clear peaks.

- **Spread.** The spread of a distribution refers to the variability of the data. If the data cluster around a single central value, the spread is smaller. The further the observations fall from the center, the greater the spread or variability of the set. (range, interquartile range, Mean Absolute Deviation, and Standard Deviation measure the spread of data)

- **Third quartile.** For a data set with median M, the third quartile is the median of the data values greater than ***M***. Example: For the data set {2, 3, 6, 7, 10, 12, 14, 15, 22, 120}, the third quartile is 15

- **Trend.** A change (either positive, negative or constant) in data values over time.

- **Two-Frequency Table.** A useful tool for examining relationships between categorical variables. The entries in the cells of a two-way table can be frequency counts or relative frequencies.

Math Class

Name________________________________ Date___________________

Standards for Mathematical Practice

1. Make sense of problems and persevere in solving them.
2. Attend to precision

Common Core State Standards
MCC9-12.S.ID. 1 Represent data with plots on the real number line (dot plots, histograms, and box plots).
MCC9-12.S.ID. 2 Use statistics appropriate to the shape of the data distribution to compare center (median, mean) and spread (interquartile range, mean absolute deviation) of two or more different data sets.
MCC9-12.S.ID. 3 Interpret differences in shape, center, and spread in the context of the data sets, accounting for possible effects of extreme data points (outliers).

Mr. Turner has two Math 2 classes. With one class, he lectured and the students took notes. In the other class, the students worked in small groups to solve math problems. After the first test, Mr. Turner recorded the student grades to determine if his different styles of teaching might have impacted student learning.

Class 1: 80, 81, 81, 75, 70, 72, 74, 76, 77, 77, 77, 79, 84, 88, 90, 86, 80, 80, 78, 82

Class 2: 70, 90, 88, 89, 86, 86, 86, 86, 84, 82, 77, 79, 84, 84, 84, 86, 87, 88, 88, 88

Analyze his student grades by calculating the mean, median, mean absolute deviation, and interquartile range. Which class do you think was the lecture and which was the small group? Why?

Draw graphs to easily compare the shapes of the distributions.

Which measure of center and spread is more appropriate to use? Explain.

Learning Task: If the Shoe Fits!

Name_______________________________ Date__________________

Common Core State Standards for Mathematical Practice

1. Make sense of problems and persevere in solving them.
2. Reason abstractly and quantitatively.
3. Construct viable arguments and critique the reasoning of others.
4. Model with mathematics.
5. Use appropriate tools strategically.

Common Core GPS:

MCC9-12.S.ID. 1 Represent data with plots on the real number line (dot plots, histograms, and box plots).

MCC9-12.S.ID. 2 Use statistics appropriate to the shape of the data distribution to compare center (median, mean) and spread (interquartile range, mean absolute deviation) of two or more different data sets.

MCC9-12.S.ID. 3 Interpret differences in shape, center, and spread in the context of the data sets, accounting for possible effects of extreme data points (outliers).

MCC9-12.S.ID. 6. Represent data on two quantitative variables on a scatter plot, and describe how the variables are related.

Welcome to CSI at School! Over the weekend, a student entered the school grounds without permission. Even though it appears that the culprit was just looking for a quiet place to study undisturbed by friends, school administrators are anxious to identify the offender and have asked for your help. The only available evidence is a suspicious footprint outside the library door. After the incident, school administrators arranged for the data in the table below to be obtained from a random sample of this high school's students. The table shows the shoe print length (in cm), height (in inches), and gender for each individual in the sample.

Shoe Print Length	Height	Gender	Shoe Print Length	Height	Gender
24	71	F	24.5	68.5	F
32	74	M	22.5	59	F
27	65	F	29	74	M
26	64	F	24.5	61	F
25.5	64	F	25	66	F
30	65	M	37	72	M
31	71	M	27	67	F
29.5	67	M	32.5	70	M
29	72	F	27	66	F
25	63	F	27.5	65	F
27.5	72	F	25	62	F

25.5	64	F	31	69	M
27	67	F	32	72	M
31	69	M	27.4	67	F
26	64	F	30	71	M
27	67	F	25	67	F
28	67	F	26.5	65.5	F
26.5	64	F	30	70	F
22.5	61	F	31	66	F
			27.25	67	F

1. Explain why this study was an observational study and not an experiment.

2. Why do you think the school's administrators chose to collect data on a random sample of students from the school? What benefit might a random sample offer?

3. Suggest a graph that might be used to use to compare the shoe print length data distributions for females and males.

4. Describe one advantage of using comparative box plots instead of comparative dot plots to display these data.

5. For each gender calculate the five-number summary for the shoe print lengths. Additionally, for each gender, determine if there are any outlying shoe print length values.

6. Construct comparative box plots for the shoe print lengths of males and females. Discuss the similarities and differences in the shoe print length distributions for the males and females in this sample.

7**.** For each gender calculate the mean shoe print length. What information does the mean shoe print length provide?

8. The mean will give us an indication of a typical shoe print length. In addition to knowing a typical length we would also like to know how much variability to expect around this length. For each gender calculate the Range; Interquartile Range; and Mean Absolute Deviation of the shoe print lengths. Interpret each of the calculated values.

9. If the length of a student's shoe print was 32 cm, would you think that the print was made by a male or a female? How sure are you that you are correct? Explain your reasoning. Use results from Questions 5 through 8 in your explanation.

10. How would you answer Question 9 if the suspect's shoe print length was 27 cm?

11. Construct a scatter plot of height (vertical scale) versus shoe print length horizontal scale) using different colors or different plotting symbols to represent the data for males and females.

(a) Interpret the scatter plot. Does it look like there is a linear relationship between height and shoe print length? Explain.

(b) Does it look like the same straight line could be used to summarize the relationship between shoe print length and height for both males and females? Explain.

(c) Based on the scatter plot, if a student's shoe print length was 30 cm, approximately what height would you predict for the person who made the shoe print? Explain how you arrived at your prediction.

Performance Task: The Basketball Star

Name________________________________ Date__________________________

Common Core State Standards for Mathematical Practice

1. Make sense of problems and persevere in solving them.
1. Model with mathematics.
2. Attend to precision.
3. Look for and make use of structure.
4. Look for and express regularity in repeated reasoning.

Common Core GPS:
MCC9-12.S.ID.1 Represent data with plots on the real number line (dot plots, histograms, and box plots).

- Choose appropriate graphs to be consistent with numerical data: dot plots, histograms, and box plots.

MCC9-12.S.ID.2 Use statistics appropriate to the shape of the data distribution to compare center (median, mean) and spread (interquartile range, standard deviation-Advanced Algebra) of two or more different data sets.

- Include review of Mean Absolute Deviation as a measure of variation.

MCC9-12.S.ID.3 Interpret differences in shape, center, and spread in the context of the data sets, accounting for possible effects of extreme data points (outliers).

- Students will examine graphical representations to determine if data are symmetric, skewed left, or skewed right and how the shape of the data affects descriptive statistics.

Is Bob or Alan a Basketball Star?

Bob's Points per Game

8, 15, 10, 10, 10, 15, 7, 8, 10, 9, 12, 11, 11, 13, 7, 8, 9, 9, 8, 10, 11, 14, 11, 10, 9, 12, 14, 14, 12, 13, 5, 13, 9, 11, 12, 13, 10, 8, 7, 8

1. Bob believes he is a basketball star and so does his friend Alan. Create a dot plot and box plot of Bob's points for the last 40 games.

2. Describe Bob's data in terms of center, spread, and shape.

3. Bob's friend Alan has the following points:

Alan's Points per Game

1, 3, 0, 2, 4, 5, 7, 7, 8, 10, 4, 4, 3, 2, 5, 6, 6, 6, 8, 8, 10, 11, 11, 10, 12, 12, 5, 6, 8, 9, 10, 15, 10, 12, 11, 11, 6, 7, 7, 8

Create a histogram of both Bob's and Alan's data.

4. Use summary statistics to compare Bob and Alan's points per game.

5. Which graphical representation best displayed Bob's and Alan's data?

6. Based on the summary statistics is either friend a basketball star? Justify your answer.

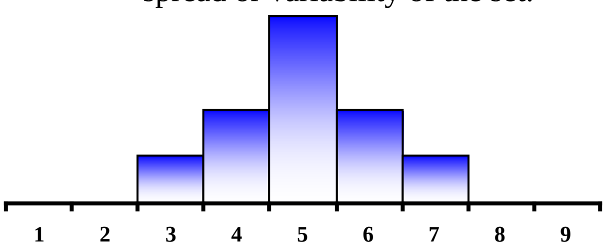
<u>Scaffolding Handout: How to Compare Distributions</u>

When you compare two or more data sets, focus on four features:

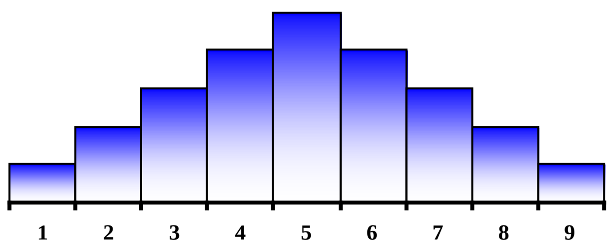
- **Center.** Graphically, the center of a distribution is the point where about half of the observations are on either side.
- **Spread**. The spread of a distribution refers to the variability of the data. If the observations cover a wide range, the spread is larger. If the observations are clustered around a single value, the spread is smaller.
- **Shape**. The shape of a distribution is described by symmetry, skewness, number of peaks, etc.
- **Unusual features**. Unusual features refer to gaps (areas of the distribution where there are no observations) and outliers.

SPREAD

The spread of a distribution refers to the variability of the data. If the data cluster around a single central value, the spread is smaller. The further the observations fall from the center, the greater the spread or variability of the set.



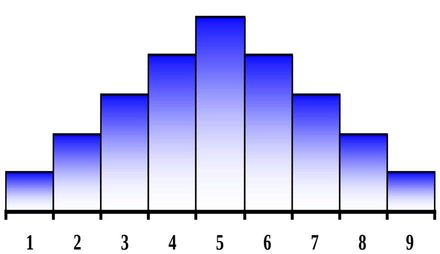
Less Spread



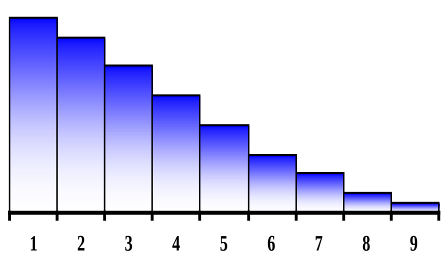
More Spread

SHAPE

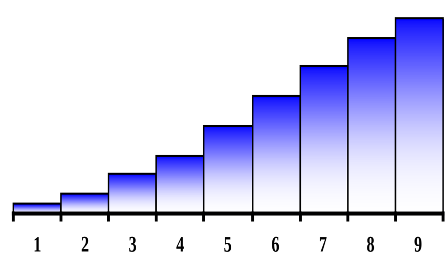
The shape of a distribution is described by symmetry, number of peaks, direction of skew, or uniformity



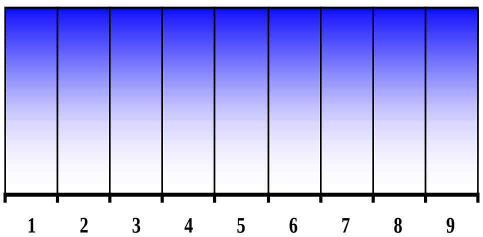
symmetric, unimodal, bell-shaped



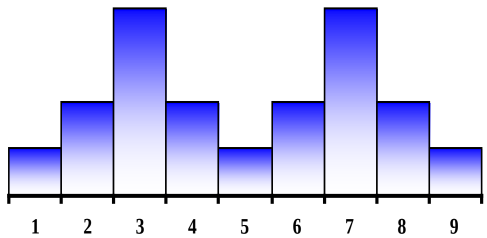
skewed right



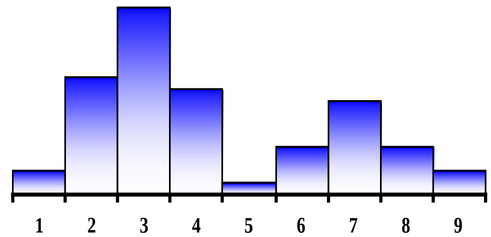
skewed left



uniform



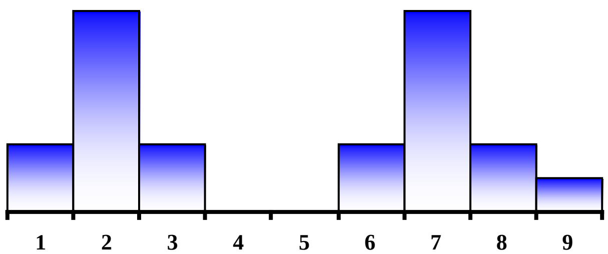
symmetric, bimodal



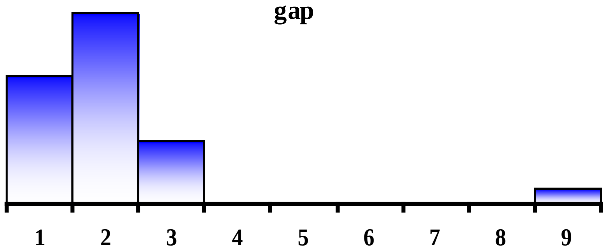
non-symmetric, bimodal

UNUSUAL FEATURES

Sometimes, statisticians refer to unusual features in a set of data. The two most common unusual features are gaps and outliers.



gap



outlier

Homework Task: Public Opinions

Name______________________ Date_______________

Common Core State Standards for Mathematical Practice
1. Make sense of problems and persevere in solving them.
2. Attend to precision.

MCC9-12.S.ID.5 Summarize categorical data for two categories in two-way frequency tables. Interpret relative frequencies in the context of the data (including joint, marginal, and conditional relative frequencies). Recognize possible associations and trends in the data.

A public opinion survey explored the relationship between age and support for increasing the minimum wage. The results are found in the following two-way frequency table

	For	Against	No Opinion	TOTAL
Ages 21-40	25	20	5	50
Ages 41-60	30	30	15	75
Over 60	50	20	5	75
TOTAL	105	70	25	200

Frequency Count

1. In the 41 to 60 age group, what percentage supports increasing the minimum wage? Explain how you arrived at your percentage.

2. What are the marginal frequencies?

3. What are the joint frequencies?

4. Why are joint and marginal frequencies important when describing trends or associations in data?

References
Jordan-Granite Consortium (2012). http://secmathccss.wordpress.com/

Homework Task: Leisure Time

1. Make sense of problems and persevere in solving them.
2. Attend to precision.

MCC9-12.S.ID.5 Summarize categorical data for two categories in two-way frequency tables. Interpret relative frequencies in the context of the data (including joint, marginal, and conditional relative frequencies). Recognize possible associations and trends in the data.

1. Using the table below, construct a table displaying conditional frequencies or the conditional distribution.

	Dance	Sports	Movies	TOTAL
Women	16	6	8	**30**
Men	2	10	8	**20**
TOTAL	**18**	**16**	**16**	50

2. After the basketball game, the statistician did not have time to compute Jana's relative frequency. Complete the table determining the relative frequency for Jana. Discuss any trends or associations from the table below concerning points scored by two basketball players.

Point Value	*Frequency for Jana*	*Relative Frequency for Jana*	*Frequency for Jill*	*Relative Frequency for Jill*
0	0		1	0.025
1	0		1	0.025
2	0		2	0.05
3	0		2	0.05
4	0		3	0.075
5	1		3	0.075
6	0		5	0.125
7	3		4	0.1
8	6		5	0.125
9	5		1	0.025
10	7		4	0.1
11	5		5	0.125
12	4		3	0.075
13	4		0	0
14	3		0	0
15	2		1	0.025
TOTALS	**40**	**1**	**40**	**1**

References

Jordan-Granite Consortium (2012). http://secmathccss.wordpress.com/

Group Learning Task: Spaghetti Regression

Adapted from: *http://txcc.sedl.org/events/previous/092806/10ApplyingStrategies/math-teks-alg1.pdf*

Common Core State Standards for Mathematical Practice

1. Make sense of problems and persevere in solving them.
2. Reason abstractly and quantitatively.
3. Construct viable arguments and critique the reasoning of others.
4. Model with mathematics.
5. Use appropriate tools strategically.
6. Attend to precision.
7. Look for and make use of structure.

Common Core State Standards

MCC9-12.S.ID.6 Represent data on two quantitative variables on a scatter plot, and describe how the variables are related.

MCC9-12.S.ID.6b Informally assess the fit of a function by plotting and analyzing residuals.

MCC9-12.S.ID.6c Fit a linear function for a scatter plot that suggests a linear association.

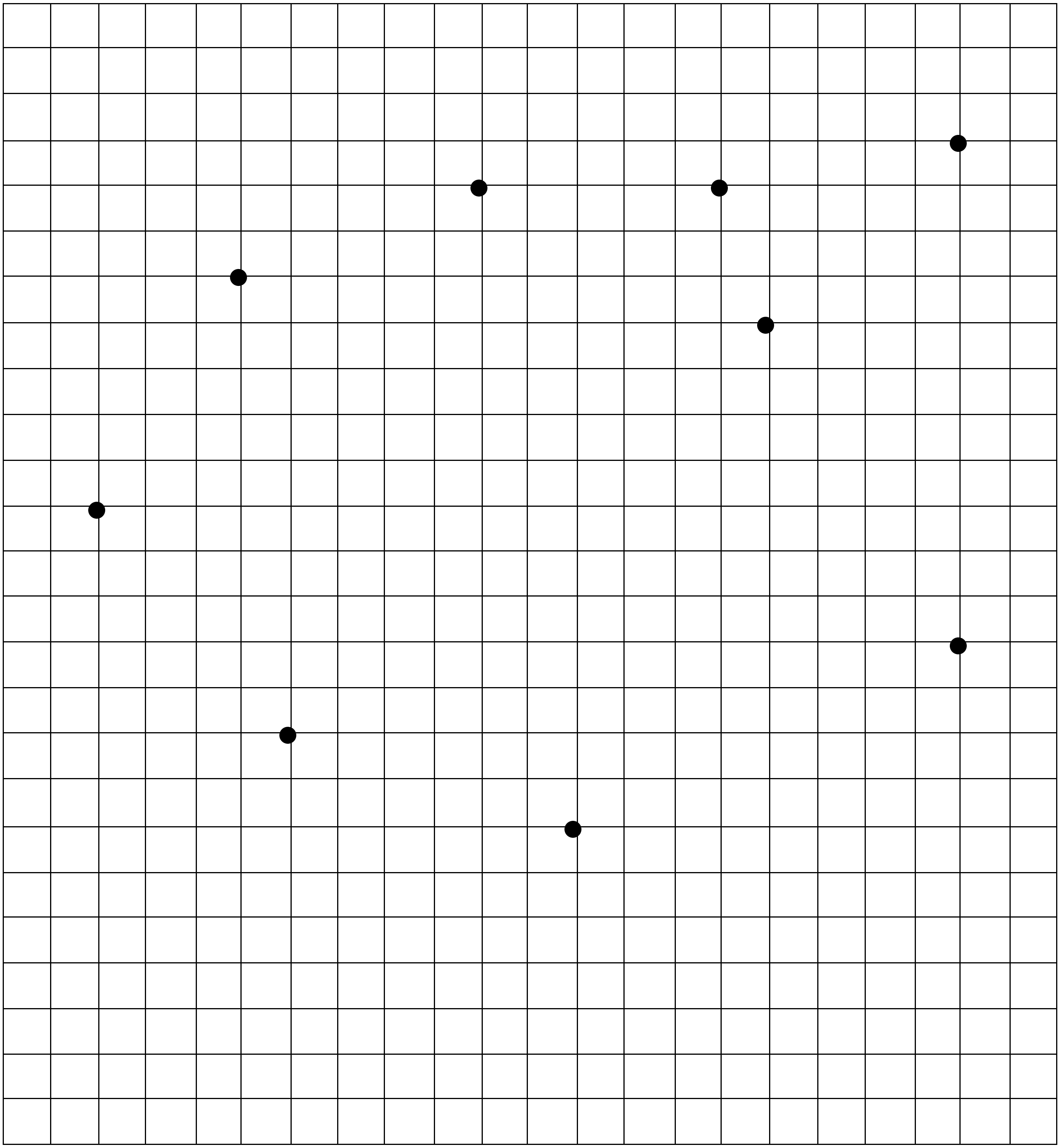
1. Examine the plot provided and visually determine a line of best-fit (or trend line) using a piece of spaghetti. Tape your spaghetti line onto your graph.

2. Now investigate the "goodness" of the fit. Use a second piece of spaghetti to measure the distance from the first point to the line. Break off this piece to represent that distance. Each person at the table must measure in the same way, so discuss the method you will use before starting. Repeat this for each point in the scatter plot.

3. Line up your "spaghetti distances" to determine who in your group has the "closest" fit. Determine the total error. (i.e., total distance from your line to the data.) Then replace the segments and tape them to your scatter plot.

Total error = ___________________cm (nearest tenth)

Scatter plot



Guided Learning Task: TV/Test Grades

Name__________________________________ **Date**____________________

Common Core State Standards for Mathematical Practice
1. Make sense of problems and persevere in solving them.
2. Reason abstractly and quantitatively.
3. Construct viable arguments and critique the reasoning of others.
4. Model with mathematics.
5. Use appropriate tools strategically.

Common Core State Standards
MCC9-12.S.ID.6 Represent data on two quantitative variables on a scatter plot, and describe how the variables are related.

MCC9-12.S.ID.6a Fit a function to the data; use functions fitted to data to solve problems in the context of the data. Use given functions or choose a function suggested by the context. Emphasize linear and exponential models.

MCC9-12.S.ID.6b Informally assess the fit of a function by plotting and analyzing residuals.

MCC9-12.S.ID.6c Fit a linear function for a scatter plot that suggests a linear association.

Interpret linear models

MCC9-12.S.ID.7 Interpret the slope (rate of change) and the intercept (constant term) of a linear model in the context of the data.

MCC9-12.S.ID.8 Compute (using technology) and interpret the correlation coefficient of a linear fit.

MCC9-12.S.ID.9 Distinguish between correlation and causation.

1. Students in Ms. Garth's Algebra II class wanted to see if there are correlations between test scores and height and between test scores and time spent watching television. Before the students began collecting data, Ms. Garth asked them to predict what the data would reveal. Answer the following questions that Ms. Garth asked her class.

   a. Do you think students' heights will be correlated to their test grades? If you think a correlation will be found, will it be a positive or negative correlation? Will it be a strong or weak correlation?

b. Do you think the average number of hours students watch television per week will be correlated to their test grades? If you think a correlation will be found, will it be a positive or negative correlation? Will it be a strong or weak correlation? Do watching TV and low test grades have a cause and effect relationship?

2. The students then created a table in which they recorded each student's height, average number of hours per week spent watching television (measured over a four-week period), and scores on two tests. Use the actual data collected by the students in Ms. Garth's class, as shown in the table below, to answer the following questions.

Student	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**
Height (in inches)	60	65	51	76	66	72	59	58	70	67	65	71	58
TV hrs/week (average)	30	12	30	20	10	20	15	12	15	11	16	20	19
Test 1	60	80	65	85	100	78	75	95	75	90	90	80	75
Test 2	70	85	75	85	100	88	85	90	90	90	95	85	85

a. Which pairs of variables seem to have a positive correlation? Explain.

b. Which pairs of variables seem to have a negative correlation? Explain.

c. Which pairs of variables seem to have no correlation? Explain.

3. For each pair of variables listed below, create a scatter plot with the first variable shown on the *y*-axis and the second variable on the *x*-axis. Are the two variables correlated positively, correlated negatively, or not correlated? Determine whether each scatter plot suggests a linear trend.

a. Score on test 1 versus hours watching television
b. Height versus hours watching television
c. Score on test 1 versus score on test 2
d. Hours watching television versus score on test 2

4. Using the statistical functions of your graphing calculator, determine a line of good fit for each scatter plot that suggests a linear trend.

Performance Task: Equal Salaries for Equal Work!

Name______________________________ Date__________________

Common Core State Standards for Mathematical Practice
1. Make sense of problems and persevere in solving them.
2. Reason abstractly and quantitatively.
3. Construct viable arguments and critique the reasoning of others.
4. Model with mathematics.
5. Use appropriate tools strategically.
6. Attend to precision.
7. Look for and make use of structure.
8. Look for and express regularity in repeated reasoning.

Common Core GPS
MCC9-12.S.ID.6 Represent data on two quantitative variables on a scatter plot, and describe how the variables are related.

MCC9-12.S.ID.6a Fit a function to the data; use functions fitted to data to solve problems in the context of the data. Use given functions or choose a function suggested by the context. Emphasize linear and exponential models.

MCC9-12.S.ID.6b Informally assess the fit of a function by plotting and analyzing residuals.

MCC9-12.S.ID.6c Fit a linear function for a scatter plot that suggests a linear association.

Interpret linear models

MCC9-12.S.ID.7 Interpret the slope (rate of change) and the intercept (constant term) of a linear model in the context of the data.

MCC9-12.S.ID.8 Compute (using technology) and interpret the correlation coefficient of a linear fit.

MCC9-12.S.ID.9 Distinguish between correlation and causation.

The data table shows the annual median earnings for female and male workers in the United States from 1984 to 2004. Use the data table to complete the task. Answer all questions in depth to show your understanding of the standards.

Year	Women's median earnings (in dollars)	Men's median earnings (in dollars)
1984	8675	17026
1985	9328	17779
1986	10016	18782
1987	10619	19818
1988	11096	20612
1989	11736	21376
1990	12250	21522
1991	12884	21857
1992	13527	21903
1993	13896	22443
1994	14323	23656
1995	15322	25018
1996	16028	25785
1997	16716	26843
1998	17716	28755
1999	18440	30079
2000	20267	30951
2001	20851	31364
2002	21429	31647
2003	22004	32048
2004	22256	32483

Data provided by U.S. Census Bureau

1. Create two scatter plots, one for women's median earnings over time and one for men's median earnings over time. Describe two things you notice about the scatter plots.

2. Terry and Tomas are trying to decide what type of model will most accurately represent the data. Terry thinks that a linear model might be most appropriate for each scatter plot. Help Terry find reasonable linear function rules for each scatter plot. Explain how you found these.

3. Using the linear models, will women's annual median earnings ever equal those of men? Why or why not?

4. Tomas thinks that an exponential model might be most appropriate for each scatter plot. Help Tomas find reasonable exponential function rules for each scatter plot. Explain how you found these.

5. Using the exponential models, will women's annual median earnings ever equal those of men? Why or why not?

6. If you answered yes to either question 3 or question 5, use that model to determine the first year women will have higher median earnings than men. Explain how you found your answer.

7. For each year listed in the table, find the ratio of women's to men's annual median earnings expressed as a percentage. Use the data to create a scatter plot of percentage versus year. Based on this graph, do you think women's annual median earnings will ever equal those of men? Why or why not?

8. Considering the results of the scatter plot in question 7 above, do you think the linear model or exponential model makes more sense? Why?

Data on earnings by gender provided by:

U.S. Census Bureau. "Table P-41. Work Experience—All Workers by Median Earnings and Sex: 1967 to2005." *Historical Income Tables—People.* www.census.gov/hhes/www/income/histinc/p41ar.html. (Date retrieved: July 24, 2007.)